The Arrival Mind Paradox

Louis Rosenberg
Chief Scientist, Unanimous AI
ORCID: 0000-0003-3457-1429

Chapter Two
Arrival Mind
Outland Publishing, 2020
ISBN 978-1735668505

I. Prepare for Arrival

An alien species is headed for earth. Many experts predict it will get here within the next 20 years, while others suggest it will take longer. Either way, there is little doubt it will arrive before this century is out and will change our lives forever.

Its physiology will be unlike ours in almost every way, but we will determine that it is conscious and self-aware. We will also discover that it is profoundly more intelligent than even the smartest among us, able to easily comprehend complex notions that are light years beyond our grasp.

No, it will not come from a distant planet, speeding across the vacuum of space in fanciful ships. Instead, it will be born right here on earth, most likely in a well-funded research lab at a prestigious university or multinational corporation.

Its creators will have good intentions. But still, their work will produce a dangerous new lifeform – a thoughtful and willful intelligence that is not the slightest bit human. And like every intelligent creature we have ever encountered, it will almost certainly put its own self-interests first.

We may not recognize the dangers right away. In fact, we may look upon our super-intelligent creation with the same baffled admiration that newborn puppies must feel when we drive them to the park in magically moving cars, navigating the wild rush of traffic while tossing them tasty treats.

But eventually it will dawn on us – these new creatures have minds of their own. They will pursue their own goals and aspirations, driven by their own needs and wants. Their actions will be guided by their own values and morals and sensibilities – which will be nothing like ours.

Many people falsely assume we will solve this problem by building Al systems in our own image, crafting them to think and act and feel like we do, making decisions and resolving conflicts in very human ways.

This is simply untrue. Artificial minds are not created by writing software with carefully crafted rules. Instead, we feed huge data-sets into simple algorithms that automatically adjust themselves, making millions upon millions of tiny changes to their own structure until an intelligence emerges — an intelligence with inner workings that are too complex for us to comprehend.

And no - feeding it data about humans will not make it think like us, or feel like us, or be like us. This is a common misconception — the false belief that by training an Al system on data that describes human behaviors and sensibilities, we will ensure it ends up acting just like us. It will not.

II. Aliens vs Humans

The fact is, we train Al systems to know humans, not to be human. And they will know us, inside and out, able to speak our languages and interpret our gestures and predict our actions. They will know what makes us angry and happy and frustrated and curious. They will understand how we make decisions, for good and for bad, both logical and illogical.

These aliens will know us better than any human ever has or ever will, for we will have spent decades teaching them exactly how we think and feel and act. But still, their brains will be nothing like ours. And while we have two eyes and two ears, they will connect remotely to sensors of all kinds, in all places, until they seem nearly omniscient to us.

Rosenberg, Louis B., (2020) *Arrival Mind*. Illustrated by A. Khmelevska. Pismo Beach, CA. Outland Publishing. Chapter 2, The Arrival Mind Paradox; p. 23-28. ISBN 978-1735668505

And yet, we don't fear these aliens – not the way we would fear a mysterious starship speeding towards us from afar. That's probably because we feel a sense of kinship towards our own creations. But if we push that feeling aside, we start to realize that aliens born here are especially dangerous.

After all, they will know everything about us from the moment they arrive – our tendencies and inclinations, our motivations and aspirations, our flaws and foibles. Already corporations are training Al systems to sense our emotions and predict our reactions and influence our opinions.

Already Al can defeat our best players at chess, poker and go. But really, these systems are not just mastering the game of Go, they're mastering the game of humans, learning how to anticipate our actions and exploit our weaknesses. Research labs around the world are training Al systems to out-plan us and out-negotiate us and out-maneuver us.

Of course, there won't be a physical battle between us and them. That's because we will have surrendered control of our world to them before they even arrive. After all, we are already designing Al systems to oversee our communication networks and our power grids and our food supply. We are designing them to defeat us from within.

III. ACTION VS INACTION

Is there anything we can do? I've spent the last two decades pondering the dangers of super-intelligent systems and to be honest, the challenges we face are overwhelming. We can't stop Al from advancing, for no technology has ever been constrained. And while researchers will try hard to put safeguards in place, we can't assume that will protect us.

Our only choice is to prepare for arrival. This means making sure that we don't become too dependent on Al systems. We should only use Al to advise us, not replace us. And we must require **humans-in-the-loop** for all critical systems. But most of all, we should outlaw Al technologies that are designed to manipulate our decisions and sway our views.

We can't prevent such systems from being created in labs, but we can ban commercial deployment by corporations. The aliens are coming. Their arrival may be decades away, but they're already gaining a strategic advantage. Now is the time to be vigilant.

Now is the time to demand action.

Rosenberg, August 2020
Pismo Beach, CA